JC13 Rec'd PCT/PTO 2 6 MAR 2001

FORM PTO-1390 (REV. 11-2000) U S DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER: 017002-012720US

U.S. APPLICATION NO. (If known, see 37 CFR 1.5: 09/806193

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US99/22259	September 24, 1999	September 25, 1998

TITLE OF INVENTION
METHOD AND APPARATUS FOR THREE-DIMENSIONAL AUDIO DISPLAY

APPLICANT(S) FOR DO/EO/US
JEAN-MARC JOT; SCOTT WARDLE

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.
4. ☒ The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau).
   b. ☐ has been communicated by the International Bureau.
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
   a. ☐ is attached hereto.
   b. ☐ has been previously submitted under 35 U.S.C. 154(d)(4).
7. ☒ Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☒ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English lanugage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☒ A FIRST preliminary amendment.
14. ☐ A SECOND or SUBSEQUENT preliminary amendment.
15. ☐ A substitute specification.
16. ☐ A change of power of attorney and/or address letter.
17. ☐ A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
18. ☐ A second copy of the published international application under 35 U.S.C. 154(d)(4).
19. ☐ A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
20. ☒ Other items or information:

    Copy of face sheet of published PCT application

JC03 Rec'd PCT/PTO 26 MAR 2001

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) 09/806193	INTERNATIONAL APPLICATION NO. PCT/US99/22259	ATTORNEY'S DOCKET NUMBER 017002-012720US

	CALCULATIONS	PTO USE ONLY
21. [X] The following fees are submitted: **BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):** Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO .... **$1000.00** International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO .... **$860.00** International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO .... **$710.00** International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) .... **$690.00** International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) **$100.00**		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 860	
Surcharge of **$130.00** for furnishing the oath or declaration later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	$	
Total claims	49 -20 =	+29	x **$18.00**	$ 522	
Independent claims	5 -3 =	+ 2	x **$80.00**	$ 160	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ **$270.00**	$	
TOTAL OF ABOVE CALCULATIONS =				$1,542	
[ ] Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2. +				$	
SUBTOTAL =				$	
Processing fee of **$130.00** for furnishing the English translation later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$	
TOTAL NATIONAL FEE =				$ 1,542	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$	
TOTAL FEES ENCLOSED =				$1,542	
				Amount to be refunded:	$
				charged:	$

a. [ ] A check in the amount of $ ______ to cover the above fees is enclosed.

b. [X] Please charge my Deposit Account No. 20-1430 in the amount of $ 1,542 to cover the above fees. A duplicate copy of this sheet is enclosed.

c. [X] The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 20-1430. A duplicate copy of this sheet is enclosed.

d. [ ] Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

George B.F. Yee
Townsend and Townsend and Crew LLP
Two Embarcadero Center, 8th fl.
San Francisco, CA 94111

SIGNATURE

J. Georg Seka
NAME

24,491
REGISTRATION NUMBER

PATENT

Attorney Docket No. 017002-012720US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re U.S. National Phase of
PCT/US99/22259 of:

JEAN-MARC JOT, et al.

Application No.: Not yet assigned

Filed: Herewith

For: METHOD AND APPARATUS FOR THREE-DIMENSIONAL AUDIO DISPLAY

PRELIMINARY AMENDMENT

San Francisco, CA 94111
March 26, 2001

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Simultaneously with the filing of this application, please amend it as indicated below. Marked-up versions of the changes to the claims are attached to this Preliminary Amendment.

IN THE CLAIMS:

Please substitute the following amended, clean versions of the indicated claims:

36. (amended) The method according to claim 34 wherein one or more of the spatial functions have their principal direction corresponding to the direction of one of the loudspeakers.

37. (amended) The method according to claim 33 including performing cross-talk cancellation of the left and right audio signals before feeding the loudspeakers.

38. (amended) The method of claim 34 further including:

producing left-front and left-back signals based on the left-channel audio signal;

producing right-front and right-back signals based on the right-channel audio signal; and

combining the left-front, left-back, right-front, and right-back signals to produce outputs suitablefor playback with a pair of front speakers and a pair of rear speakers.

REMARKS:

Claims 1-49 are pending.

Amendment is made to eliminate all multiple dependencies from the claims, thereby avoiding the need to pay the multiple dependent surcharge.

Respectfully submitted,

J. Georg Seka
Reg. No. 24,491

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: (415) 576-0200
Fax: (415) 576-0300
JGS:tp
SF 1204666 v1

MARKED-UP VERSION OF THE CHANGES TO THE CLAIMS

IN THE CLAIMS:

36. (amended) The method according to [claims 34 or 35] claim 34 wherein one or more of the spatial functions have their principal direction corresponding to the direction of one of the loudspeakers.

37. (amended) The method according to [claims 33 or 36] claim 33 including performing cross-talk cancellation of the left and right audio signals before feeding the loudspeakers.

38. (amended) The method of [claims 34 or 35] claim 34 further including:

producing left-front and left-back signals based on the left-channel audio signal;

producing right-front and right-back signals based on the right-channel audio signal; and

combining the left-front, left-back, right-front, and right-back signals to produce outputs suitablefor playback with a pair of front speakers and a pair of rear speakers.

METHOD AND APPARATUS FOR THREE-DIMENSIONAL AUDIO DISPLAY

FIELD OF THE INVENTION

The present invention relates generally to audio recording, and more specifically to the mixing, recording and playback of audio signals for reproducing real or virtual three-dimensional sound scenes at the eardrums of a listener using loudspeakers or headphones.

BACKGROUND

A well-known technique for artificially positioning a sound in a multi-channel loudspeaker playback system consists of weighting an audio signal by a set of amplifiers feeding each loudspeaker individually. This method, described e. g. in [Chowning71], is often referred to as "discrete amplitude panning" when only the loudspeakers closest to the target direction are assigned non-zero weights, as illustrated by the graph of panning functions in Fig. 1. Although Fig. 1 shows a two-dimensional loudspeaker layout, the method can be extended with no difficulty to three-dimensional loudspeaker layouts, as described e. g. in [Pulkki97]. A drawback of this technique is that it requires a high number of channels to provide a faithful reproduction of all directions. Another drawback is that the geometrical layout of the loudspeakers must be known at the encoding and mixing stage.

An alternative approach, described in [Gerzon85], consists of producing a 'B-Format' multi-channel signal and reproducing this signal over loudspeakers via an 'Ambisonic' decoder, as illustrated in Fig. 2. Instead of discrete panning functions, the B Format uses real-valued spherical harmonics. The zero-order spherical harmonic function is named W, while the three first-order harmonics are denoted X, Y, and Z. These functions are defined as follows:

$$W(\theta, \varphi) = 1$$

$$X(\theta, \varphi) = \cos(\varphi)\cos(\theta)$$

$$Y(\theta, \varphi) = \cos(\varphi)\sin(\theta)$$

$$Z(\theta, \varphi) = \sin(\varphi)$$

where θ and φ denote respectively the azimuth and elevation angles of the sound source with respect to the listener, expressed in radians. An advantage of this technique over the discrete panning method is that B Format encoding does not require knowledge of the loudspeaker layout, which is taken into account in the design of the decoder. A second advantage is that a real-world B-Format recording can be produced with practical microphone technology, known as the 'Soundfield Microphone' [Farrah79]. As illustrated in Fig. 2, this allows for combining microphone-encoded sounds with electronically encoded sounds to produce a single B-format recording. First-order Ambisonic decoders do not reconstruct the acoustic pressure information at the ears of the listener except at low frequencies (below about 700 Hz). As described e. g. in [Bamford95], the frequency range can be extended by increasing the order of spherical harmonics, but only at the expense of a higher number of encoding channels and loudspeakers.

3-D audio reproduction techniques which specifically aim at reproducing the acoustic pressure at the two ears of a listener are usually termed binaural techniques. This approach is illustrated in Fig. 3 and reviewed e. g. in [Jot95]. A binaural recording can be produced by inserting miniature microphones in the ear canals of an individual or dummy head. Binaural encoding of an audio signal (also called binaural synthesis) can be performed by applying to a sound signal a pair of left and right filters modeling the head-related transfer functions (HRTFs) measured on an individual or a dummy head for a given direction. As shown in Fig. 3, a HRTF can be modeled as a cascaded combination of a delaying element and a minimum-phase filter, for each of the left and right channels. A binaurally encoded or recorded signal is suitable for playback over headphones. For playback over loudspeakers, a cross-talk canceller is used, as described e. g. in [Gardner97].

Conventional binaural techniques can provide a more convincing 3-D audio reproduction, over headphones or loudspeakers, than the previously described techniques. However, they are not without their own drawbacks and difficulties.

- Compared to discrete amplitude panning or B-Format encoding, binaural synthesis involves a significantly larger amount of computation for each sound source. An accurate finite impulse response (FIR) model of an HRTF typically requires a 1-ms long response, i. e. approximately 100 additions and multiplies per sample

period at a sample rate of 48 kHz, which amounts to 5 MIPS (million instructions per second).

- The HRTF can only be measured at a set of discrete positions around the head. Designing a binaural synthesis system which can faithfully reproduce any direction and smooth dynamic movements of sounds is a challenging problem involving interpolation techniques and time-variant filters, implying an additional computational effort.

- The binaurally recorded or encoded signal contains features related to the morphology of the torso, head, and pinnae. Therefore the fidelity of the reproduction is compromised if the listener's head is not identical to the head used in the recording or the HRTF measurements. In headphone playback, this can cause artifacts such as an artificial elevation of the sound, front-back confusions or inside-the-head localization.

- In reproduction over two loudspeakers, the listener must be located at a specific position for lateral sound locations to be convincingly reproduced (beyond the azimuth of the loudspeakers), while rear or elevated sound locations cannot be reproduced reliably.

[Travis96] describes a method for reducing the computational cost of the binaural synthesis and addresses the interpolation and dynamic issues. This method consists of combining a panning technique designed for N-channel loudspeaker playback and a set of N static binaural synthesis filter pairs to simulate N fixed directions (or "virtual loudspeakers") for playback over headphones. This technique leads to the topology of Fig. 4a, where a bank of binaural synthesis filters is applied after panning and mixing of the source signals. An alternative approach, described in [Gehring96], consists of applying the binaural synthesis filters before panning and mixing, as illustrated in Fig. 4b. The filtered signals can be produced off-line and stored so that only the panning and mixing computations need to be performed in real time. In terms of reproduction fidelity, these two approaches are equivalent. Both suffer from the inherent limitations of the multi-channel positioning techniques. Namely, they require a large number of encoding channels to faithfully reproduce the localization and timbre of sound signals in any direction.

[Lowe95] describes a variation of the topology of Fig. 4a, in which the directional encoder generates a set of two-channel (left and right) audio signals, with a direction-dependent time delay introduced between the left and right channels, and each two-channel signal is panned between front, back and side "azimuth placement" filters.[Chen96] uses an analysis method known as principal component analysis (PCA) to model any set of HRTFs as a weighted sum of frequency-dependent functions weighted by functions of direction. The two sets of functions are listener-specific (uniquely associated to the head on which the HRTF were measured) and can be used to model the left filter and the right filter applied to the source signal in the directional encoder. [Abel97] also shows the topologies of Figs. 4a and 4b and uses a singular value decomposition (SVD) technique to model a set of HRTFs in a manner essentially equivalent to the method described in [Chen96], resulting in the simultaneous solution for a set of filters and the directional panning functions.

There remains a need for a computationally efficient technique for high-fidelity 3-D audio encoding and mixing of multiple audio signals. It is desirable to provide an encoding technique that produces a non listener-specific format. There is a need for a practical recording technique and suitably designed decoders to provide faithful reproduction of the pressure signals at the ears of a listener, over headphones or two-channel and multi-channel loudspeaker playback systems.

SUMMARY OF THE INVENTION

A method for positioning an audio signal includes selecting a set of spatial functions and providing a set of amplifiers. The gains of the amplifiers being dependent on scaling factors associated with the spatial functions. An audio signal is received and a direction for the audio signal is determined. The scaling factors are adjusted depending on the direction. The amplifiers are applied to the audio signal to produce first encoded signals. The audio signal is then delayed. The second filters are then applied to the delayed signal to produce second encoded signals. The resulting encoded signals contain directional information. In one embodiment of the invention, the spatial functions are the spherical harmonic functions. The spherical harmonics may include zero-order and first-order harmonics and higher order harmonics. In another embodiment, the spatial functions include discrete panning functions.

Further in accordance with the method of the invention, a decoding of the directionally encoded audio includes providing a set of filters. The filters are defined based on the selected spatial functions.

An audio recording apparatus includes first and second multiplier circuits having adjustable gains. A source of an audio signal is provided, the audio signal having a time-varying direction associated therewith. The gains are adjusted based on the direction for the audio. A delay element inserts a delay into the audio signal. The audio and delayed audio are processed by the multiplier circuits, thereby creating directionally encoded signals. In one embodiment, an audio recording system comprises a pair of soundfield microphones for recording an audio source. The soundfield microphones are spaced apart at the positions of the ears of a notional listener.

According to the invention, a method for decoding includes deriving a set of spectral functions from preselected spatial functions. The resulting spectral functions are the basis for digital filters which comprise the decoder.

According to the invention, a decoder is provided comprising digital filters. The filters are defined based on the spatial functions selected for the encoding of the audio signal. The filters are arranged to produce output signals suitable for feeding into loudspeakers.

The present invention provides an efficient method for 3-D audio encoding and playback of multiple sound sources based on the linear decomposition of HRTF using spatial panning functions and spectral functions, which

- guarantees accurate reproduction of ITD cues for all sources over the whole frequency range
- uses predetermined panning functions.

The use of predetermined panning functions offers the following advantages over methods of the prior art which use principal components analysis or singular value decomposition to determine panning functions and spectral functions:

- efficient implementation in hardware or software
- non-individual encoding/recording format
- adaptation of the decoder to the listener
- improved multi-channel loudspeaker playback

Two particularly advantageous choices for the panning functions are detailed, offering additional benefits:

- Spherical harmonics
- allow to make recordings using available microphone technology (a pair of Soundfield microphones)
- yield a recording format that is a superset of the B format standard
- associated to a special decoding technique for multi-channel loudspeaker playback
- Discrete panning functions
- guarantees exact reproduction of chosen directions
- increased efficiency of implementation (by minimizing the number of non-zero panning weights for each source)
- associated to a special decoding technique for multi-channel loudspeaker playback

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Discrete panning over 4 loudspeakers. Example of discrete panning functions.

Figure 2: B-format encoding and recording. Playback over 6 loudspeakers using Ambisonic decoding.

Figure 3: Binaural encoding and recording. Playback over 2 speakers using cross-talk cancellation.

Figure 4: (a) Post-filtering topology. (b) Pre-filtering topology.

Figure 5: (a) Post-filtering and (b) pre-filtering topologies, with control of interaural time difference for each sound source.

Figure 6: Binaural B Format encoding with decoding for playback over over headphones.

Figure 7: Original and reconstructed HRTF with Binaural B Format (first-order reconstruction).

Figure 8: Binaural B Format reconstruction filters (amplitude frequency response).

Figure 9: Binaural B Format decoder for playback over 4 speakers.

Figure 10: Binaural Discrete Panning using 6 encoding channels, with decoder for playback over 2 speakers with cross-talk cancellation.

Figure 11: Binaural Discrete Panning using 6 encoding channels, with decoder for playback over 4 speakers with cross-talk cancellation.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Modeling HRTF using predetermined spatial functions

Given a set of N spatial panning functions $\{g_i(\theta, \varphi), i = 0, 1, \ldots N\text{-}1\}$ the procedure for modeling HRTF according to the present invention is as follows. This procedure is associated to the topologies described in Fig. 5a and Fig. 5b for directionnally encoding one or several audio signals and decoding them for playback over headphones.

1. Measuring HRTFs for a set of positions $\{(\theta_p, \varphi_p), p = 1, 2, \ldots P\}$. The sets of left-ear and right-ear HRTFs will be denoted, respectively, as:

   $\{L(\theta_p, \varphi_p, f)\}$ and $\{R(\theta_p, \varphi_p, f)\}$, for $p = 1, 2, \ldots P$, where f denotes frequency.

2. Extracting the left and right delays $t_L(\theta_p, \varphi_p)$ and $t_R(\theta_p, \varphi_p)$ for every position. Denoting $T(\theta, \varphi, f) = \exp(2\pi j\, f\, t(\theta, \varphi))$, the time-delay operator of duration t, expressed in the frequency domain, the left-ear and right-ear HRTFs are expressed by:

   $$L(\theta_p, \varphi_p, f) = T_L(\theta_p, \varphi_p, f)\, \underline{L}(\theta_p, \varphi_p, f),$$

   $$R(\theta_p, \varphi_p, f) = T_R(\theta_p, \varphi_p, f)\, \underline{R}(\theta_p, \varphi_p, f), \text{ for } p = 1, 2, \ldots P.$$

3. Equalization removing a common transfer function from all HRTFs measured on one ear. This transfer function can include the effect of the measuring apparatus, loudspeaker, and microphones used. It can also be the delay-free HRTF $\underline{L}$ (or $\underline{R}$) measured for one particular direction (free-field equalization), or a transfer function representing an average of all the delay-free HRTFs $\underline{L}$ (or $\underline{R}$) measured over all positions (diffuse-field equalization).

4. Symmetrization, whereby the HRTFs and the delays are corrected in order to verify the natural left-right symmetry relations:

   $\underline{R}(\theta, \varphi, f) = \underline{L}(2\pi-\theta, \varphi, f)$ and $t_L(\theta, \varphi) = t_R(2\pi-\theta, \varphi)$.

5. Derivation of the set of reconstruction filters $\{L_i(f)\}$ and $\{R_i(f)\}$ satisfying the approximate equations:

   $\underline{L}(\theta_p, \varphi_p, f) \cong \sum_{\{i=0,\dots N-1\}} g_i(\theta_p, \varphi_p)\, L_i(f)$,

   $\underline{R}(\theta_p, \varphi_p, f) \cong \sum_{\{i=0,\dots N-1\}} g_i(\theta_p, \varphi_p)\, R_i(f)$, for $p = 1, 2, \dots P$.

In practice, the measured HRTFs are obtained in the digital domain. Each HRTF is represented as a complex frequency response sampled at a given number of frequencies over a limited frequency range, or, equivalently, as a temporal impulse response sampled at a given sample rate. The HRTF set $\{\underline{L}(\theta_p, \varphi_p, f)\}$ or $\{\underline{R}(\theta_p, \varphi_p, f)\}$ is represented, in the above decomposition, as a complex function of frequency in which every sample is a function of the spatial variables θ and φ, and this function is represented as a weighted combination of the spatial functions $g_i(\theta, \varphi)$. As a result, a sampled complex function of frequency is associated to each spatial function $g_i(\theta, \varphi)$, which defines the sampled frequency response of the corresponding filter $L_i(f)$ or $R_i(f)$. It is noted that, due to the linearity of the Fourier transform, an equivalent decomposition would be obtained if the the frequency variable f were replaced by the time variable in order to reconstruct the time-domain representation of the HRTF.

The equalization and the symmetrization of the HRTF sets $\underline{L}(\theta_p, \varphi_p, f)$ and $\underline{R}(\theta_p, \varphi_p, f)$, are not necessary to carrying out the invention. However, performing these operations eliminates some of the artifacts associated to the HRTF measurement method. Thus, it may be preferable to perform these operations for their practical advantages.

Step 2 is optional and is associated to the binaural synthesis topologies described in Figs. 5a and 5b, where the delays $t_L(\theta, \varphi)$ and $t_R(\theta, \varphi)$ are introduced in the directional encoding module for each sound source. If step 2 is not applied, the binaural synthesis topologies of Figs. 4a and 4b can be used. If the delay extraction procedure is appropriately performed (as discussed below) the topologies of Figs. 5a and 5b will provide a higher fidelity with fewer encoding channels. It will be noted that adding or subtracting a common delay offset to $t_L(\theta, \varphi)$ and $t_R(\theta, \varphi)$ in the encoding module will have no effect over the perceived direction of sounds during playback, even if the delay offset varies with direction, as long as the interaural time delay difference (ITD), defined below, is preserved for each direction.

$$ITD(\theta, \varphi) = t_R(\theta, \varphi) - t_L(\theta, \varphi).$$

It is noted that the above procedure differs from the methods of the prior art. Conventional analytical techniques, such as PCA and SVD, simultaneously produce the spectral functions and the spatial functions which minimize the least-squares error between the original HRTFs and the reconstructed HRTFs for a given number of channels N. In the elaboration of the present invention, it is recognized in particular, that these earlier methods suffer from the following drawbacks:

- The spatial panning functions cannot be chosen *a priori.*
- The choice of error criterion to be minimized (mean squared error) enables the resolution of the approximation problem via tractable linear algebra. However, the technique does not guarantee that the model of the HRTF thus obtained is optimal in terms of perceived reproduction for a given number of encoding channels.

In comparison, the technique in accordance with the present invention permits *a priori* selection of the spatial functions, from which the spectral functions are derived. As will be apparent from the following description, several benefits of the present invention will result from the possibility of choosing the panning functions *a priori* and from using a variety of techniques to derive the associated reconstruction filters.

An immediate advantage of the invention is that the encoding format in which sounds are mixed in Fig. 5a is devoid of listener specific features. As discussed below, it is possible, without causing major degradations in reproduction fidelity, to use a listener-independent model of the ITD in carrying out the invention.

Generally, it is possible to make a selection of spatial panning functions and tune the reconstruction filters to achieve practical advantages such as:

- enabling improved reproduction over multi-channel loudspeaker systems,
- enabling the production of microphone recordings,
- preserving a high fidelity of reproduction in chosen directions or regions of space even with a low number of channels.

Two particular choices of spatial panning functions will be detailed in this description: spherical harmonic functions and discrete panning functions. Practical methods for designing the set of reconstruction filters $L_i(f)$ and $R_i(f)$ will be described in more detail. From the discussion which follows, it will be clear to a person of ordinary skill in the relevant art that other spatial functions can be used and that alternative techniques for producing the corresponding reconstruction filters are available.

Delay extraction techniques

The extraction of the interaural time delay difference, $ITD(\theta_p, \varphi_p)$, from the HRTF pair $L(\theta_p, \varphi_p, f)$ and $R(\theta_p, \varphi_p, f)$ is performed as follows.

Any transfer function $H(f)$ can be uniquely decomposed into its all-pass component and its minimum-phase component as follows:

$$H(f) = \exp(j\,\psi(f))\, H_{min}(f)$$

where $\psi(f)$, called the excess-phase function of $H(f)$, is defined by

$$\psi(f) = \mathrm{Arg}(H(f)) - \mathrm{Re}(\mathrm{Hilbert}(-\mathrm{Log}|H(f)|)).$$

Applying this decomposition to the HRTFs $L(\theta_p, \varphi_p, f)$ and $R(\theta_p, \varphi_p, f)$, we obtain the corresponding excess-phase functions, $\psi_R(\theta_p, \varphi_p, f)$ and $\psi_L(\theta_p, \varphi_p, f)$, and the corresponding minimum-phase HRTFs, $L_{min}(\theta_p, \varphi_p, f)$ and $R_{min}(\theta_p, \varphi_p, f)$.

The interaural time delay difference, $ITD(\theta_p, \varphi_p)$, can be defined, for each direction (θ_p, φ_p), by a linear approximation of the interaural excess-phase difference:

$$\psi_R(\theta, \varphi, f) - \psi_L(\theta, \varphi, f) \cong 2\pi f\, ITD(\theta, \varphi).$$

In practice, this approximation may be replaced by various alternative methods of estimating the ITD, including time-domain methods such as methods using the cross-correlation function of the left and right HRTFs or methods using a threshold detection technique to estimate an arrival time at each ear. Another possibility is to use a formula for modeling the variation of ITD vs. direction. For instance,

- the spherical head model with diametrally opposite ears yields

$$ITD(\theta, \varphi) = r/\mathrm{c}\ [\ \arcsin(\cos(\varphi)\sin(\theta)) + \cos(\varphi)\sin(\theta)\ ],$$

- the free-field model –where the ears are represented by two points separated by the distance $2r$– yields

$$ITD(\theta, \varphi) = 2r/\mathrm{c}\ \cos(\varphi)\sin(\theta),$$

where c denotes the speed of sound. In these two formulas, the value of the radius r can be chosen so that $ITD(\theta_p, \varphi_p)$ is as large as possible without exceeding the value derived from the linear approximation of the interaural excess-phase difference. In a digital implementation, the value of $ITD(\theta_p, \varphi_p)$, can be rounded to the closest integer number of samples, or the interaural excess-phase difference may be approximated by the combination of a delay unit and a digital all-pass filter.

The delay-free HRTFs, $\underline{L}(\theta_p, \varphi_p, f)$ and $\underline{R}(\theta_p, \varphi_p, f)$, from which the reconstruction filters $L_i(f)$ and $R_i(f)$ will be derived, can be identical, respectively, to the minimum-phase HRTF $L_{min}(\theta_p, \varphi_p, f)$ and $R_{min}(\theta_p, \varphi_p, f)$.

Whatever the method used to extract or model the interaural time delay difference from the measured HRTF, it can be regarded as an approximation of the interaural excess-phase difference $\psi_R(\theta, \varphi, f) - \psi_L(\theta, \varphi, f)$ by a model function $\psi(\theta, \varphi, f)$:

$$\psi_R(\theta, \varphi, f) - \psi_L(\theta, \varphi, f) \cong \psi(\theta, \varphi, f).$$

It may be advantageous, in order to improve the fidelity of the 3-D audio reproduction according to the present invention, to correct for the error made in this phase difference approximation, by incorporating the residual excess-phase difference into the delay-free HRTFs $\underline{L}(\theta_p, \varphi_p, f)$ and $\underline{R}(\theta_p, \varphi_p, f)$ as follows:

$$\underline{L}(f) = L_{min}(f)\exp(\mathrm{j}\,\phi_L(f)) \text{ and } \underline{R}(f) = R_{min}(f)\exp(\mathrm{j}\,\phi_R(f)),$$

where $\phi_L(f)$ and $\phi_R(f)$ satisfy

$$\phi_R(f) - \phi_L(f) = \psi_R(f) - \psi_L(f) - \psi(\theta, \varphi, f),$$

and either $\phi_L(f) = 0$ or $\phi_R(f) = 0$, as appropriate to ensure that the delay-free HRTFs $\underline{L}(\theta_p, \varphi_p, f)$ and $\underline{R}(\theta_p, \varphi_p, f)$ are causal transfer functions.

Application of spherical harmonic functions for encoding and recording

General definition of spherical harmonics.

Of particular interest in the following description are the zero-order harmonic W and the first-order harmonics X, Y and Z defined earlier, as well as the second-order harmonics, U and V, and the third-order harmonics, S and T, defined below.

$$U(\theta, \varphi) = \cos^2(\varphi)\cos(2\theta)$$
$$V(\theta, \varphi) = \cos^2(\varphi)\sin(2\theta)$$
$$S(\theta, \varphi) = \cos^3(\varphi)\cos(3\theta)$$
$$T(\theta, \varphi) = \cos^3(\varphi)\sin(3\theta)$$

Advantages of spherical harmonics include:

- mathematically tractable, closed form -> interpolation between directions
- mutually orthogonal
- spatial interpretation (e. g. front-back difference)
- facilitates recording

Fig. 6 illustrates this method in the case where the minimum-phase HRTFs are decomposed over spherical harmonics limited to zero and first order. The directional encoding of the input signal producesan 8-channel encoded signal herein referred to as a “Binaural B Format” encoded signal. The mixer provides for mixing of additional source signals, including synthesized sources. Conversely, 8 filters are used to decode this format into a binaural output signal. The method can be extended to include any or all of the above higher-order spherical harmonics. Using the higher orders provides for more accurate reconstruction of HRTFs, especially at high frequencies (above 3 kHz).

As discussed above, a Soundfield microphone produces B format encoded signals. As such, a Soundfield microphone can be characterized by a set of spherical harmonic functions. Thus from Fig. 6, it can be seen that encoding a sound in accordance with the invention to produce Binaural B Format encoded signals, simulates a free-field

recording using two Soundfield microphones located at the notional position of the two ears. This simulation is exact if the directional encoder provides ITD according to the following free-field model:

$$ITD(\theta, \varphi) = t_R(\theta, \varphi) - t_L(\theta, \varphi) = d/c \cos(\varphi) \sin(\theta),$$

where d is the distance between the microphones. If the ITD model provided in the encoder takes into account the diffraction of sound around the head or a sphere, the encoded signal and the recorded signal will differ in the value of the ITD for sounds away from the median plane. This difference can be reduced, in practice, by adjusting the distance between the two microphones to be slightly larger than the distance between the two ears of the listener.

The Binaural B Format recording technique is compatible with currently existing 8-channel digital recording technology. The recording can be decoded for reproduction over headphones through the bank of 8 filters $L_i(f)$ and $R_i(f)$ shown on Fig. 6, or decoded over two or more loudspeakers using methods to be described below. Before decoding, additional sources can be encoded in Binaural B Format and mixed into the recording.

The Binaural B Format offers the additional advantage that the set of four left or right channels can be used with conventional Ambisonic decoders for loudspeaker playback. Other advantages of using spherical harmonics as the spatial panning functions in carrying out the invention will be apparent in connection to multi-channel loudspeaker playback, offering an improved fidelity of 3-D audio reproduction compared to Ambisonic techniques.

Derivation of the reconstruction filters

For clarity, the derivation of the N reconstruction filters $L_i(f)$ will be illustrated in the case where the spatial panning functions $g_i(\theta_p, \varphi_p)$ are spherical harmonics. However, the methods described are general and apply regardless of the choice of spatial functions.

The problem is to find, for a given frequency (or time) f, a set of complex scalars $L_i(f)$ so that the linear combination of the spatial functions $g_i(\theta_p, \varphi_p)$ weighted by the $L_i(f)$

approximates the spatial variation of the HRTF $\underline{L}(\theta_p, \varphi_p, f)$ at that frequency (or time). This problem can be conveniently represented by the matrix equation

$$\underline{L} = \boldsymbol{G}\, L,$$

where

- the set of HRTF $\underline{L}(\theta_p, \varphi_p, f)$ defines the $P\times 1$ vector $\underline{L}$, P being the number of spatial directions
- each spatial panning function $g_i(\theta_p, \varphi_p)$ defines the $P\times 1$ vector G_i, and the matrix $\boldsymbol{G}$ is the $P\times N$ matrix whose columns are the vectors G_i
- the set of reconstruction filters $L_i(f)$ defines the $N\times 1$ vector of unknowns L.

The solution which minimizes the energy of the error is given by the pseudo inversion

$$L = (\boldsymbol{G}^{\mathrm{T}}\, \boldsymbol{G})^{-1}\, \boldsymbol{G}^{\mathrm{T}}\, \underline{L},$$

where $(\boldsymbol{G}^{\mathrm{T}}\, \boldsymbol{G})$, known as the Gram matrix, is the $N\times N$ matrix formed by the dot products $\boldsymbol{G}(i, k) = G_i^{\mathrm{T}}\, G_k$ of the spatial vectors. The Gram matrix is diagonal if the spatial vectors are mutually orthogonal.

Simplest case: the sampled spatial functions are mutually orthogonal => filters are derived by orthogonal projection of the HRTF on the individual spatial functions (dot product computed at each frequency). Example: 2-D reproduction with regular azimuth sampling. If sampled functions are not mutually orthogonal, multiply by inverse of Gram matrix to ensure correct reconstruction.

Even when the panning functions $g_i(\theta, \varphi)$ are mutually ortogonal, as is the case with spherical harmonics, the vectors G_i obtained by sampling these functions may not be orthogonal. This happens typically if the spatial sampling is not uniform (as is often the case with 3-D HRTF measurements). This problem can be remedied by redefining the spatial dot product so as to approximate the continuous integral of the product of two spatial functions

$$< g_i, g_k > = 1/(4\pi) \int_\theta \int_\varphi\, g_i(\theta, \varphi)\, g_k(\theta, \varphi) \cos(\varphi)\, d\theta\ d\varphi$$

by

$$< g_i, g_k > = \sum_{\{p = 1,\dots P\}}\, g_i(\theta_p, \varphi_p)\, g_k(\theta_p, \varphi_p)\, dS(p) = G_i^{\mathrm{T}}\, \Delta\, G_k$$

where Δ is a diagonal $P \times P$ matrix with $\Delta(p, p) = dS(p)$, and $dS(p)$ is proportional to a notional solid angle covered by the HRTF measured for the direction (θ_p, φ_p). This definition yields the generalized pseudo inversion equation

$$L = (\boldsymbol{G}^{T} \Delta \boldsymbol{G})^{-1} \boldsymbol{G}^{T} \Delta \underline{L},$$

where the diagonal matrix Δ can be used as a spatial weighting function in order to achieve a more accurate 3-D audio reproduction in certain regions of space compared to others, and the modified Gram matrix $(\boldsymbol{G}^{T} \Delta \boldsymbol{G})$ ensures that the solution minimizes the mean squared error.

Additional possibility: project on a subset of the chosen set of spatial functions using above methods. Then project the residual error over other spatial functions (cf aes16). Example: to optimize fidelity of reconstruction in horizontal plane, project on W, X, Y first, and then project error on Z. Note that process can be iterated in more than 2 steps.

By combining the above techniques, it is possible, for a given set of spatial panning functions, to achieve control over chosen perceptual aspects of the 3-D audio reproduction, such as the front/back or up/down discrimination or the accuracy in particular regions of space.

Fig. 7 illustrates the performance of the method for reconstructing the HRTF magnitude spectra in the horizontal plane ($\varphi = 0$). For this reconstruction, only 3 channels per ear are necessary, since the Z channel is not used. The original data are diffuse-field equalized HRTFs derived from measurements on a dummy head. Due to the limitation to first-order harmonics, the reconstruction matches the original magnitude spectra reasonably well up to about 2 or 3 kHz, but the performance tends to degrade with increasing frequency. For large-scale applications, a gentle degradation at high frequencies can be acceptable, since inter-individual differences in HRTFs typically become prominent at frequencies above 5 kHz. The frequency responses of the reconstruction filters obtained in this case are shown on Fig. 8.

Adaptation of the reconstruction filters to the listener

An advantage of a recording mad in accordance with the invention over a conventional two-channel dummy head recording is that, unlike prior art encoded

signals, binaural B format encoded signals do not contain spectral HRTF features. These features are only introduced at the decoding stage by the reconstruction filters $L_i(f)$. Contrary to a conventional binaural recording, a Binaural B Format recording allows listener-specific adaptation at the reproduction stage, in order to reduce the occurrence of artifacts such as front-back reversals and in-head or elevated localization of frontal sound events.

Listener-specific adaptation can be achieved even more effectively in the context of a real-time digital mixing system. Moreover, the technique of the present invention readily lends itself to a real-time mixing approach and can be conveniently implemented as it only involves the correction of the head radius r for the synthesis of ITD cues and the adaptation of the four reconstruction filters $L_i(f)$. If diffuse-field equalization is applied to the headphones and to the measured HRTF, and therefore to the reconstruction filters $L_i(f)$, the adaptation only needs to address direction-dependent features related to the morphology of the listener, rather than variations in HRTF measurement apparatus and conditions.

Application of discrete panning functions

Definition: functions which minimize the number of non-zero panning weights for any direction: 2 weights in 2D and 3 weights in 3D. For each panning function, there is a direction where this panning function reaches unity and is the only non-zero panning function. Example given in Fig. 1 for 2D case. Many variations possible.

An advantage of discrete panning functions: fewer operations needed in encoding module (multiplying by panning weight and adding into the mix is only necessary for the encoding channels which have non-zero weights).

The projection techniques described above can be used to derive the reconstruction filters. Alternatively, it can be noted that each discrete panning function covers a particular region of space, and admits a "principal direction" (the direction for which the panning weight reaches 1). Therefore, a suitable reconstruction filter can be the HRTF corresponding to that principal direction. This will guarantee exact reconstruction of the HRTF for that particular direction. Alternatively, a combination of the principal direction and the nearest directions can be used to derive the reconstruction filter. When it is desired to design a 3D audio display system which

offers maximum fidelity for certain directions of the sound, it is straightforward to design a set of panning functions which will admit these specific directions as principal directions.

Methods for playback over loudspeakers

When used in the topologies of Figs. 5a and 5b, the set of reconstruction filters obtained according to the present invention will provide a two-channel output signal suitable for high-fidelity 3D audio playback over headphones. As illustrated in Fig. 3, this two channel signal can be further processed through a cross-talk cancellation network in order to provide a two-channel signal suitable for playback over two loudspeakers placed in front of the listener. This technique can produce convincing lateral sound images over a frontal pair of loudspeakers, covering azimuths up to about ±120°. However, lateral sound images tend to collapse into the loudspeakers in response to rotations and translations of the listener's head. The technique is also less effective for sound events assigned to rear or elevated positions, even when the listener sits at the "sweet spot".

Fig. 9 illustrates how, in the case of spherical harmonic panning functions, the reconstruction filters $L_i(f)$ can be utilized to provide improved reproduction over multi-channel loudspeaker playback systems. An advantage of the Binaural B Format is that it contains information for discriminating rear sounds from frontal sounds. This property can be exploited in order to overcome the limitations of 2-channel transaural reproduction, by decoding over a 4-channel loudspeaker setup. The 4-channel decoding network, shown in Fig. 9, makes use of the sum and difference of the W and X signals.

The binaural signal is decomposed as follows:

$$L(\theta, \varphi, f) = LF(\theta, \varphi, f) + LB(\theta, \varphi, f)$$

where LF and LB are the "front" and "back" binaural signals, defined by:

$$LF(\theta, \varphi, f) = 0.5\ \{[W(\theta, \varphi)+X(\theta, \varphi)]\ [L_W(f)+L_X(f)] + Y(\theta, \varphi)\ L_Y(f) + Z(\theta, \varphi)\ L_Z(f)\}$$

$$LB(\theta, \varphi, f) = 0.5\ \{[W(\theta, \varphi)-X(\theta, \varphi)]\ [L_W(f)-L_X(f)] + Y(\theta, \varphi)\ L_Y(f) + Z(\theta, \varphi)\ L_Z(f)\}$$

It can be verified that $LB = 0$ for $(\theta, \varphi) = (0, 0)$ and that $LF = 0$ for $(\theta, \varphi) = (\pi, 0)$. The network of Fig. 9 is designed to eliminate front-back confusions, by reproducing frontal sounds over the front loudspeakers and rear sounds over the rear loudspeakers, while elevated or lateral sounds are reproduced via both pairs of loudspeakers. This significantly improves the reproduction of lateral, rear or elevated sound images compared to a 2-channel loudspeaker setup (or to 4-channel loudspeaker reproduction using conventional pairwise amplitude panning or Ambisonic techniques). The listener is also allowed to move more freely than with 2-channel loudspeaker reproduction. By exploiting the Z component, a similar approach can be used to decode the binaural B format over a 3-D loudspeaker setup (comprising loudspeakers above or below the horizontal plane).

Fig. 11 illustrates how the present invention, applied with discrete panning functions, can be advantageously used to provide three-dimensional audio playback over two loudspeakers placed in front of the listener, with cross-talk cancellation. In this implementation of the invention, the discrete panning functions $g_1(\theta, \varphi)$ and $g_2(\theta, \varphi)$ are chosen so that their principal directions coincide, respectively, with the directions of the left and right loudspeakers from the listener's head (the principal direction of the discrete panning function $g_i(\theta, \varphi)$ is defined as (θ_i, φ_i) verifying $g_i(\theta_i, \varphi_i) = 1.0$ and $g_j(\theta_i, \varphi_i) = 0$ for $j \neq I$). Furthermore, the reconstruction filters and the cross-talk cancellation networks are free-field equalized, for each ear, with respect to the direction of the closest loudspeaker. As a result of these conditions, it can be verified that, if an audio signal is panned to the direction of one of the two loudspeakers, it is fed with no modification to that loudspeaker and cancelled out from the output feeding the other loudspeaker. Therefore, the resulting loudspeaker playback system combines, in conjunction with the previously described advantages of the present invention, the advantage of conventional discrete panning systems and the advantages of binaural reproduction techniques using cross-talk cancellation.

The following notations are used in Fig. 10 and Fig. 11:

- $\underline{L}_{i/j}$ denotes the ratio of two delay-free HRTFs:

  $$\underline{L}_{i/j} = \underline{L}(\theta_i, \varphi_i, f) \,/\, \underline{L}(\theta_j, \varphi_j, f);$$

- $L_{i/j}$ denotes the ratio of two delay-free HRTFs combined with the time difference between them:

$$L_{ij} = \exp(2\pi j f [\, t(\theta_i, \varphi_i) - t(\theta_j, \varphi_j) \,])\, \underline{L}(\theta_i, \varphi_i, f) \,/\, \underline{L}(\theta_j, \varphi_j, f).$$

Fig. 11 illustrates how the decoder of Fig. 10 can be modified to offer further improved three-dimensional audio reproduction over four loudspeakers arranged in a front pair and a rear pair. The method used is similar to the method used in the system of Fig. 9, in that a front cross-talk canceller and a rear cross-talk canceller are used, and they receive different combinations of the left and right encoded signals. These combinations are designed so that frontal sounds are reproduced over the front loudspeakers and rear sounds are reproduced over the rear loudspeakers, while elevated or lateral sounds are reproduced via both pairs of loudspeakers. Fig. 11 shows an embodiment of the present invention using 6 encoding channel for each ear, where channels 1 and 2 are front left and right channels, channels 5 and 4 are rear left and right channels, and channels 3 and 6 are lateral and/or elevated channels. A particular advantageous property of this embodiment is that, if an audio signal is panned towards the direction of one of the four loudspeakers (corresponding to the principal direction of one of the channels 1, 2, 4, or 5), it is fed with no modification to that loudspeaker and cancelled out from the output feeding the three other loudspeakers. It is noted that, generally, the systems of Fig. 10 or Fig. 11 can be extended to include larger numbers of encoding channels without departing from the principles characterizing the present invention, and that, among these encoding channels, one or more can have their principal direction outside of the horizontal plane so as to provide the reproduction of elevated sounds or of sounds located below the horizontal plane.

What is claimed is:

1. A method for positioning of an audio signal comprising steps of:

selecting a set of spatial functions, each having an associated scaling factor;

providing a first set of amplifiers and a second set of amplifiers, the gains of the amplifiers being a function of the scaling factors;

receiving a first audio signal;

providing a direction representing the direction of the source of the first audio signal;

adjusting the scaling factors depending on the direction;

applying the first set of amplifiers to the first audio signal to produce first encoded signals;

delaying the first audio signal to produce a delayed audio signal; and

applying the second set of amplifiers to the delayed audio signal to produce second encoded signals.

2. The method of claim 1 wherein the spatial functions are spherical harmonic functions.

3. The method of claim 2 wherein the spherical harmonic functions include at least the first-order harmonics.

4. The method of claim 1 wherein the spatial functions are discrete panning functions.

5. The method of claim 1 wherein for each of the first and second sets of amplifiers, the gain of each amplifier is based on the B-format encoding scheme.

6. The method of claim 1 further including:

providing a third set of amplifiers and a fourth set of amplifiers, the gains of the amplifiers being a function of the scaling factors;

receiving a second audio signal;

providing a direction representing the direction of the source of the second audio signal;

adjusting the scaling factors depending on the direction;

applying the third set of amplifiers to the second audio signal to produce third encoded signals;

delaying the second audio signal to produce a second delayed audio signal;

applying the fourth set of amplifiers to the second delayed audio signal to produce fourth encoded signals;

mixing the first and the third encoded signals, or the first and the fourth encoded signals; and

mixing the second and the fourth encoded signals, or the second and the third encoded signals.

7. The method of claim 6 wherein the second signal is a synthesized audio signal.

8. The method of claim 1 further including a decoding the encoded signals, the decoder comprising filters defined based on the spatial functions.

9. An audio recording apparatus for directionally encoding an audio signal comprising:

a source of an audio signal, the audio signal having a time-varying direction associated therewith;

a first set of multiplier circuits, each having a gain factor adaptable according to a direction for the audio signal, each having an input to receive the audio source, each having an output;

a delay element having an input coupled to the audio source and having an output; and

a second set of multiplier circuits, each having a gain factor adaptable according to a direction for the audio signal, each having an input to receive the output of the delay element, each having an output;

whereby the outputs of the first and second multiplier circuits comprise encoded audio signals.

10. The apparatus of claim 9 wherein the source includes a source of a synthesized audio signal.

11. The apparatus of claim 9 wherein the gain factors of the first and second multiplier circuits are based on spherical harmonic functions.

12. The apparatus of claim 11 wherein the spherical harmonic functions include at least zero- and first-order harmonics.

13. The apparatus of claim 9 wherein the gain factors of the first and second multiplier circuits are based on discrete panning functions.

14. The apparatus of claim 9 further including a data storage device having an interface effective for receiving and storing the outputs of the multiplier circuits.

15. A 3-dimensional audio recording system comprising:

a first soundfield microphone to produce first directionally encoded audio signals; and

a second soundfield microphone to produce second directionally encoded audio signals;

the first and second soundfield microphones are proximate each other at the positions of the ears of a notional listener;

wherein the first and second directionally encoded audio signals represent a 3-dimensional audio recording.

16. The system of claim 15 further including a storage device for storing the first and second directionally encoded audio signals.

17. The system of claim 16 further including A/D circuitry for converting outputs of the microphones to digital signals, whereby the digital signals can be stored on the storage device.

18. The system of claim 15 wherein the first and second microphones are spaced apart by a distance substantially equal to the width of a human head.

19. The system of claim 15 wherein the first and second soundfield microphones are characterized by a set of spatial functions, the system further including a decoder for receiving the first and second directionally encoded signals to produce an audio signal, the decoder comprising filters defined based on the spatial functions.

20. A method of producing an audio signal from directionally encoded audio signals comprising steps of:

receiving directionally encoded audio signals according to a set of spatial functions;

generating a set of spectral functions based on the spatial functions;

providing a first set of decoding filters defined by left spectral functions;

providing a second set of decoding filters defined by right spectral functions;

applying the first decoding filters to the encoded audio signals to produce a left-channel audio signal; and

applying the second decoding filters to the encoded audio signals to produce a right-channel audio signal.

21. The method of claim 20 wherein the set of spatial functions is defined by $\{g_i(\theta, \varphi), i = 0, 1, \ldots N\text{-}1\}$ and the step of generating the spectral functions includes providing $L_i(f)$ and $R_i(f)$ such that $\sum_{\{i = 0, \ldots N-1\}} g_i(\theta_p, \varphi_p)\, L_i(f)$ approximates $\underline{L}(\theta_p, \varphi_p, f)$ and $\sum_{\{i = 0, \ldots N-1\}} g_i(\theta_p, \varphi_p)\, R_i(f)$ approximates $\underline{R}(\theta_p, \varphi_p, f)$, where $\underline{L}(\theta_p, \varphi_p, f)$ is a set of left-ear HRTFs and $\underline{R}(\theta_p, \varphi_p, f)$ is a set of right-ear HRTFs, where $\{(\theta_p, \varphi_p), p = 1, 2, \ldots P\}$ is a set of directions and f is frequency.

22. The method of claim 21 wherein $\underline{L}(\theta_p, \varphi_p, f)$ and $\underline{R}(\theta_p, \varphi_p, f)$ are delay-free HRTFs.

23. The method of claim 21 wherein providing $L_i(f)$ includes solving, at each frequency f, the vector equation $\underline{L} \cong \boldsymbol{G} L_i$, where:

the set of left-ear HRTFs $\underline{L}(\theta_p, \varphi_p, f)$ define a P×1 vector $\underline{L}$,

$\boldsymbol{G}$ is a $P \times N$ matrix whose columns are $P \times 1$ vectors G_i, $i = 0, 1, \ldots N\text{-}1$

each of the N spatial functions $g_i(\theta_p, \varphi_p)$ defines the vector G_i, and

the set of $L_i(f)$ defines the $N \times 1$ vector L.

24. The method of claim 23 wherein providing $L_i(f)$ is obtained by $L = (\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\underline{L}$.

25. The method of claim 24 wherein providing $L_i(f)$ includes projecting a $P \times 1$ vector $\underline{L}$ formed by the set of left-ear HRTFs $\underline{L}(\theta_p, \varphi_p, f)$ over each of $P \times 1$ vectors G_i formed by the spatial functions $g_i(\theta_p, \varphi_p)$ to compute the scalar product L_i.

26. The method according to claim 25 wherein an $N \times 1$ vector L formed by the scalar products L_i is multiplied by the inverse of the Gram matrix $\boldsymbol{G}^T\boldsymbol{G}$.

27. The method of claim 23 wherein providing $L_i(f)$ is obtained by $L = (\boldsymbol{G}^T \Delta \boldsymbol{G})^{-1}\boldsymbol{G}^T\Delta\underline{L}$, where Δ is a diagonal $P \times P$ matrix where the P diagonal elements are weights applied to the individual directions (θ_p, φ_p), $p = 1, 2, \ldots P$.

28. The method of claim 20 where each weight is proportional to a solid angle associated with the corresponding direction.

29. The method of claim 28 wherein the spatial functions are spherical harmonic functions.

30. The method of claim 21 wherein the spherical harmonic functions include at least zero- and first-order harmonics.

31. The method of claim 20 wherein the spectral functions define filters $L_W(f)$, $L_X(f)$, $L_Y(f)$, and $L_Z(f)$, effective for decoding B-format encoded signals W_L, X_L, Y_L, Z_L, W_R, X_R, Y_R, and Z_R, wherein the left-channel audio signal is defined by $W_L\,L_W(f) + X_L\,L_X(f) + Y_L\,L_Y(f) + Z_L\,L_Z(f)$ and the right-channel audio signal is defined by $W_R\,L_W(f) + X_R\,L_X(f)$ -

$Y_R\, L_Y(f) + Z_R\, L_Z(f)$; whereby the left- and right-channel audio signals are suitable for playback with headphones.

32. The method of claim 20 wherein the spectral functions define filters $L_W(f)$, $L_X(f)$, $L_Y(f)$, and $L_Z(f)$ effective for decoding B-format encoded signals W_L, X_L, Y_L, Z_L, W_R, X_R, Y_R, and Z_R; wherein the left-channel audio signal comprises two signals

a first signal $LF = 0.5\, \{[W_L + X_L]\, [L_W(f) + L_X(f)] + Y_L\, L_Y(f) + Z_L\, L_Z(f)\}$ and

a second signal $LB = 0.5\, \{[W_L - X_L]\, [L_W(f) - L_X(f)] + Y_L\, L_Y(f) + Z_L\, L_Z(f)\}$;

and wherein the right-channel audio signal comprises two signals

a first signal $RF = 0.5\, \{[W_R + X_R]\, [L_W(f) + L_X(f)] + Y_R\, L_Y(f) + Z_R\, L_Z(f)\}$ and

a second signal $RB = 0.5\, \{[W_R - X_R]\, [L_W(f) - L_X(f)] - Y_R\, L_Y(f) + Z_R\, L_Z(f)\}$;

whereby the left- and right- channel audio signals are suitable for playback over a pair of front speakers and a pair of rear speakers.

33. The method of claim 32 further including:

performing a first cross-talk cancellation on the *LF* and *RF* signals to feed the front speakers; and

performing a second cross-talk cancellation on the *LB* and *RB* signals to feed the rear speakers.

34. The method of claim 20 wherein the spatial functions are discrete panning functions having a direction, called a principal direction, where the spatial function is maximum and wherein all other spatial functions are zero.

35. The method of claim 34 wherein the spectral function associated with each spatial function is the delay-free HRTF for the corresponding principal direction.

36. The method according to claims 34 or 35 wherein one or more of the spatial functions have their principal direction corresponding to the direction of one of the loudspeakers.

37. The method according to claims 33 or 36 including performing cross-talk cancellation of the left and right audio signals before feeding the loudspeakers.

38. The method of claims 34 or 35 further including:

producing left-front and left-back signals based on the left-channel audio signal;

producing right-front and right-back signals based on the right-channel audio signal; and

combining the left-front, left-back, right-front, and right-back signals to produce outputs suitable for playback with a pair of front speakers and a pair of rear speakers.

39. The method of claim 38 further including:

performing a first cross-talk cancellation on the left-front and right-front signals to feed the front speakers; and

performing a second cross-talk cancellation on the left-back and right-back signals to feed the rear speakers.

40. The method of claim 39 wherein one or more of the spatial functions have their principal direction corresponding to the direction of the loudspeakers.

41. A method for reproducing an audio scene comprising:

selecting set of spatial functions;

producing directionally encoded audio signals including receiving a first audio source and applying the spatial functions to the first audio source to produce first encoded signals; and

decoding the encoded audio signals, including generating spectral functions based on the first spatial functions and applying the spectral functions to the encoded audio signals.

42. The method of claim 41 further including delaying the first audio source to produce a delayed source, applying the spatial functions to the delayed source to produce second encoded signals, the first and second signals comprising directionally encoded audio signals.

43. The method of claim 41 wherein the step of producing directionally encoded audio signals further includes receiving a second audio source, applying the spatial functions to the second audio source to produce second encoded signals, and mixing the first and second encoded signals.

44. The method of claim 43 wherein the second audio source is a synthesized audio signal.

45. The method of claim 41 wherein the spatial functions are spherical harmonic functions.

46. The method of claims 45 wherein the spherical harmonic functions include at least zero- and first-order harmonics.

47. The method of claim 41 wherein the spatial functions are discrete panning functions.

48. The method of claim 41 wherein the step of applying the spectral functions to the directionally encoded audio signals includes providing a set of filters defined by the spectral functions and feeding the encoded audio signals into the filters to produce reconstructed audio signals.

49. The method of claim 41 further including performing a cross-talk cancellation operation on the reconstructed audio signals to produce output suitable for playback with speakers.

FIG. 1

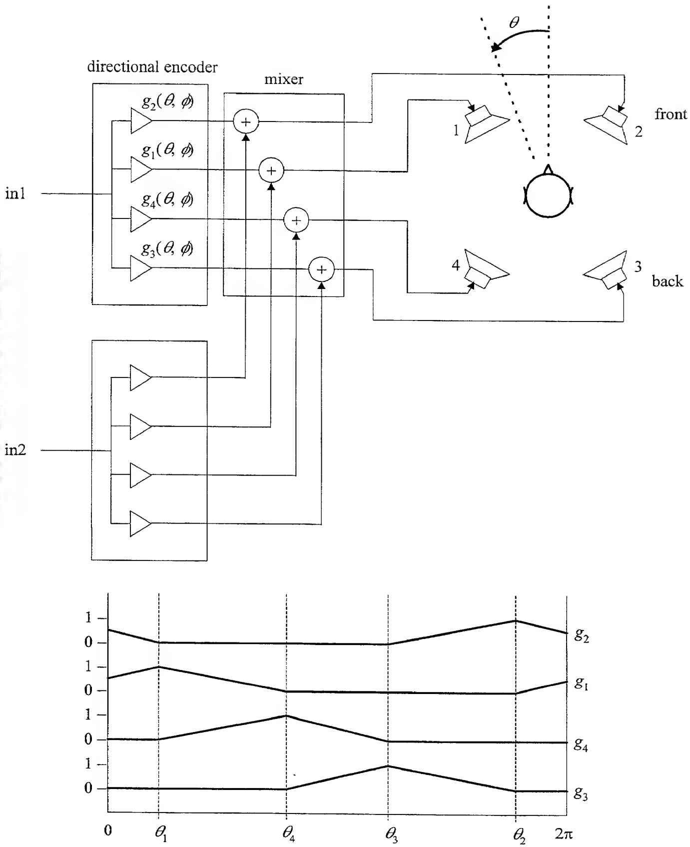

directional encoder
mixer
$g_2(\theta, \phi)$
$g_1(\theta, \phi)$
$g_4(\theta, \phi)$
$g_3(\theta, \phi)$
in1
in2
θ
1
2
front
4
3
back
1
0
1
0
1
0
1
0
g_2
g_1
g_4
g_3
0
θ_1
θ_4
θ_3
θ_2
2π

FIG. 2

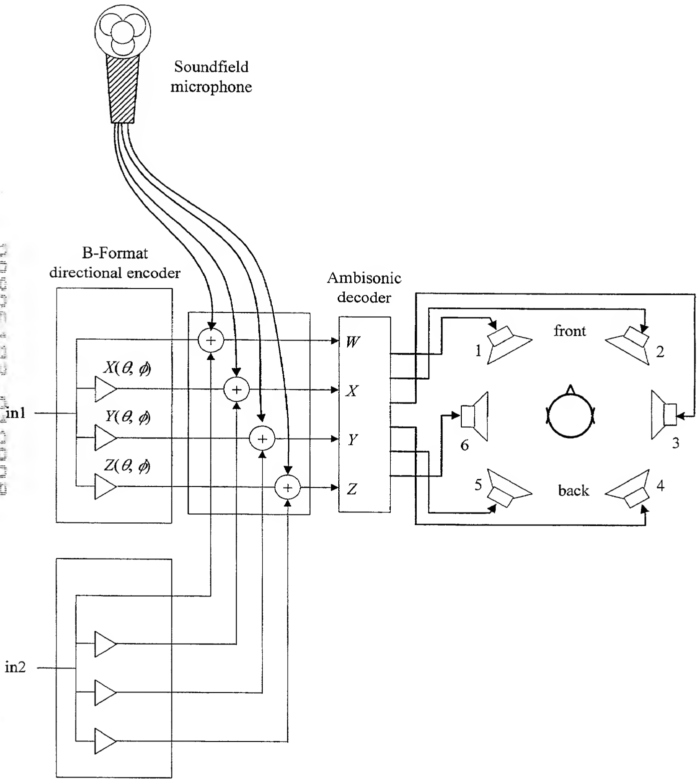

Soundfield
microphone
B-Format
directional encoder
X(θ, φ)
Y(θ, φ)
Z(θ, φ)
in1
in2
Ambisonic
decoder
W
X
Y
Z
front
back
1
2
3
4
5
6

FIG. 3

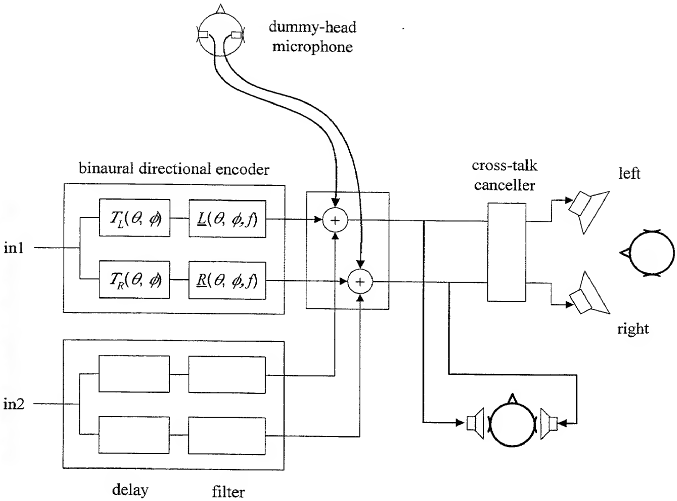
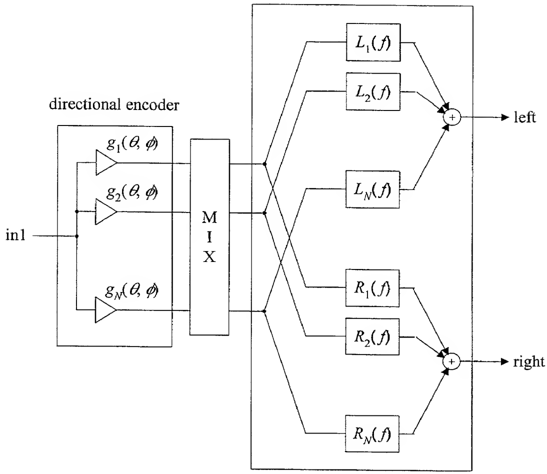

dummy-head
microphone
binaural directional encoder
$T_L(\theta, \phi)$
$\underline{L}(\theta, \phi, f)$
$T_R(\theta, \phi)$
$\underline{R}(\theta, \phi, f)$
in1
in2
delay
filter
cross-talk
canceller
left
right

FIG. 4(a)



reconstruction filter bank
(headphone decoder)

FIG. 4(b)

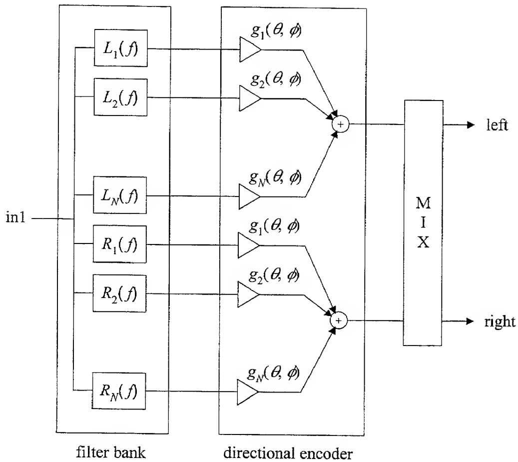

$L_1(f)$
$L_2(f)$
$L_N(f)$
$R_1(f)$
$R_2(f)$
$R_N(f)$
$g_1(\theta, \phi)$
$g_2(\theta, \phi)$
$g_N(\theta, \phi)$
$g_1(\theta, \phi)$
$g_2(\theta, \phi)$
$g_N(\theta, \phi)$
in1
+
+
M
I
X
left
right
filter bank
directional encoder

FIG. 5(a)

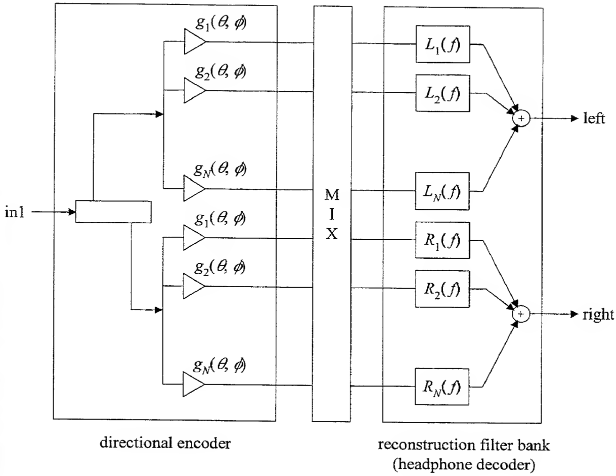

$g_1(\theta, \phi)$
$g_2(\theta, \phi)$
$g_N(\theta, \phi)$
in1
$g_1(\theta, \phi)$
$g_2(\theta, \phi)$
$g_N(\theta, \phi)$
M
I
X
$L_1(f)$
$L_2(f)$
$L_N(f)$
$R_1(f)$
$R_2(f)$
$R_N(f)$
+
+
left
right
directional encoder
reconstruction filter bank
(headphone decoder)

FIG. 5(b)

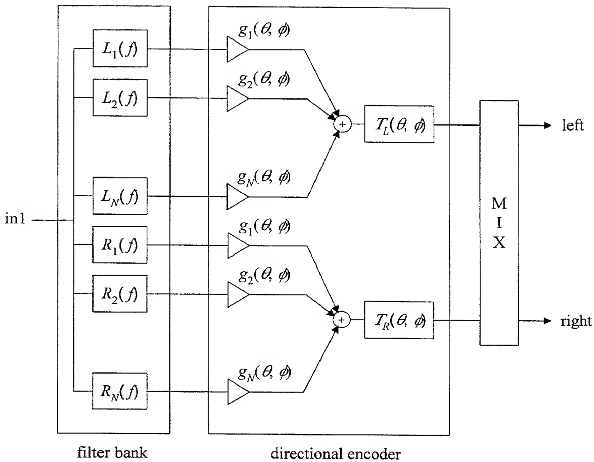

$L_1(f)$
$L_2(f)$
$L_N(f)$
in1
$R_1(f)$
$R_2(f)$
$R_N(f)$
$g_1(\theta, \phi)$
$g_2(\theta, \phi)$
$g_N(\theta, \phi)$
$g_1(\theta, \phi)$
$g_2(\theta, \phi)$
$g_N(\theta, \phi)$
+
+
$T_L(\theta, \phi)$
$T_R(\theta, \phi)$
M
I
X
left
right
filter bank
directional encoder

FIG. 6

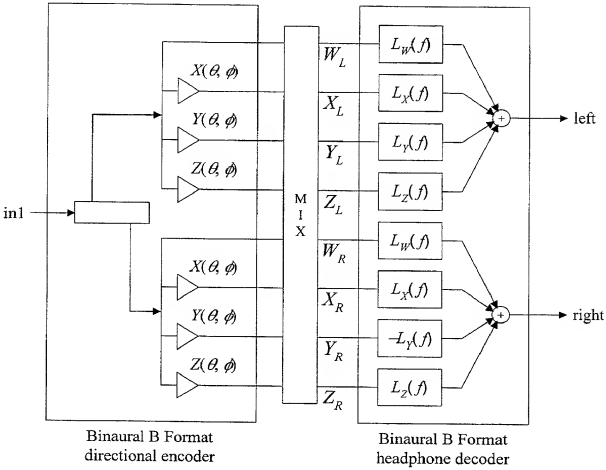

in1
$X(\theta, \phi)$
$Y(\theta, \phi)$
$Z(\theta, \phi)$
$X(\theta, \phi)$
$Y(\theta, \phi)$
$Z(\theta, \phi)$
M I X
W_L
X_L
Y_L
Z_L
W_R
X_R
Y_R
Z_R
$L_W(f)$
$L_X(f)$
$L_Y(f)$
$L_Z(f)$
$L_W(f)$
$L_X(f)$
$-L_Y(f)$
$L_Z(f)$
left
right
Binaural B Format directional encoder
Binaural B Format headphone decoder

FIG. 7

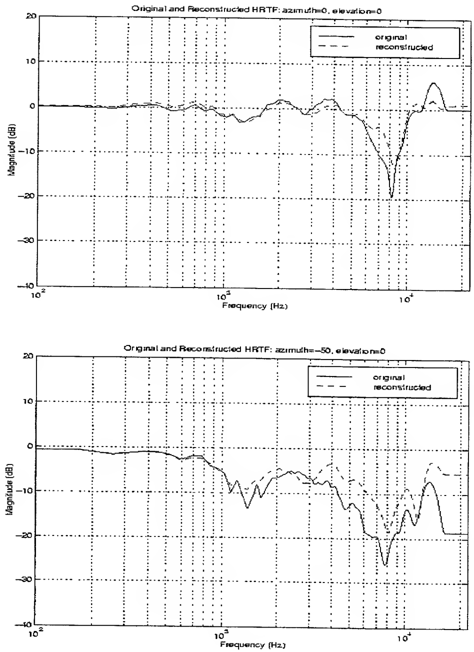

Original and Reconstructed HRTF: azimuth=0, elevation=0
original
reconstructed
Magnitude (dB)
20
10
0
−10
−20
−30
−40
10^2
10^3
10^4
Frequency (Hz)
Original and Reconstructed HRTF: azimuth=−50, elevation=0
original
reconstructed
Magnitude (dB)
20
10
0
−10
−20
−30
−40
10^2
10^3
10^4
Frequency (Hz)

FIG. 8

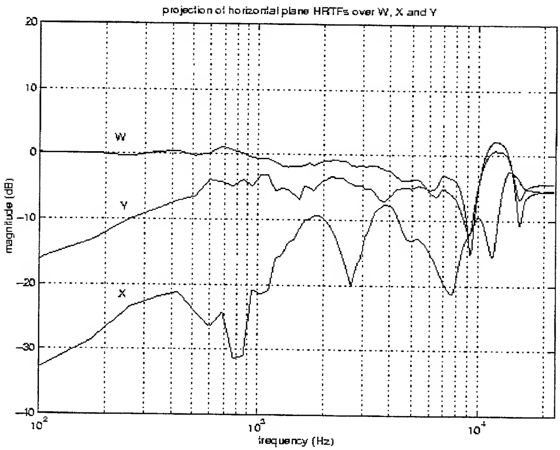
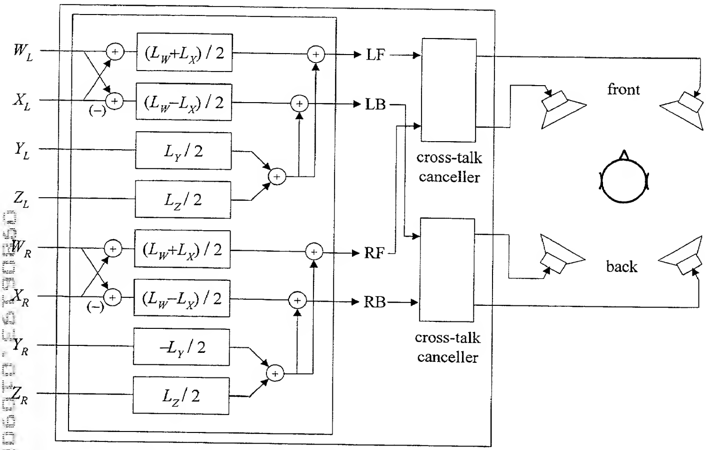

projection of horizontal plane HRTFs over W, X and Y
20
10
0
-10
-20
-30
-40
magnitude (dB)
W
Y
X
10^2
10^3
10^4
frequency (Hz)

FIG. 9



4-channel Binaural B Format louspeaker decoder

FIG. 10

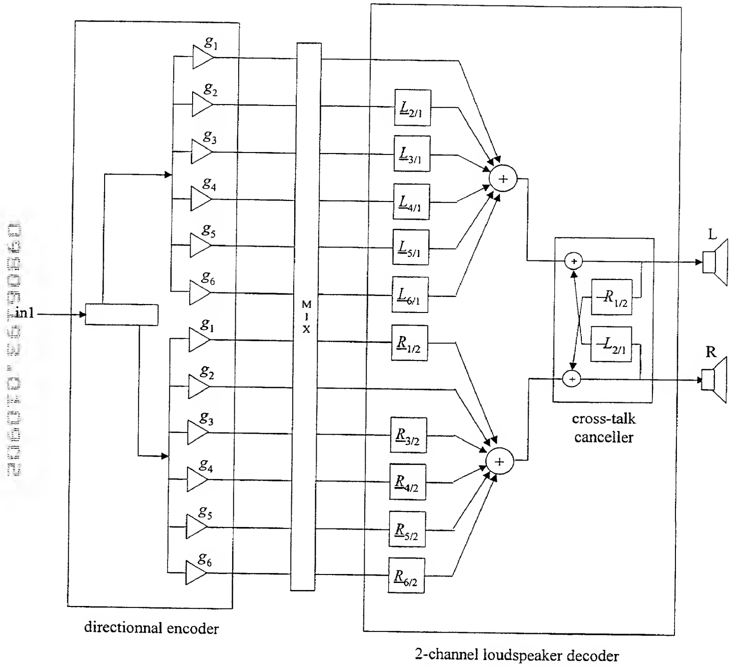
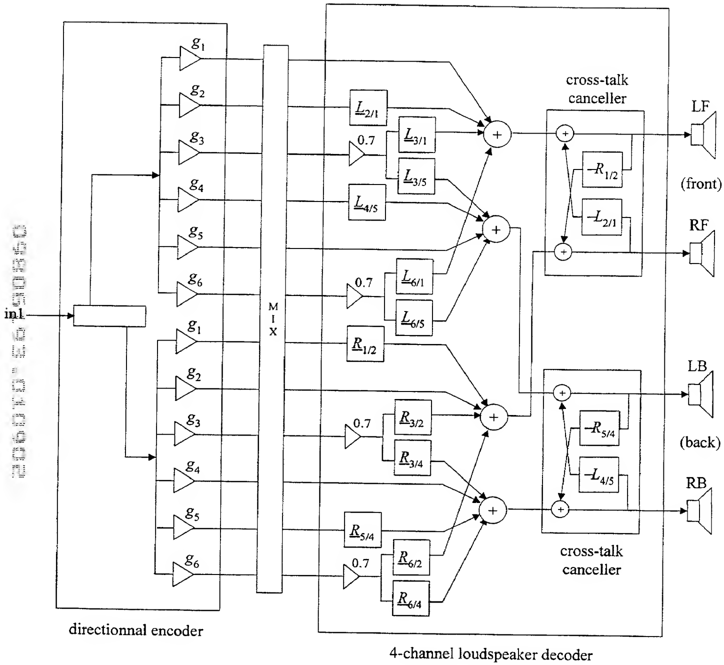

in1
g_1
g_2
g_3
g_4
g_5
g_6
g_1
g_2
g_3
g_4
g_5
g_6
M I X
$L_{2/1}$
$L_{3/1}$
$L_{4/1}$
$L_{5/1}$
$L_{6/1}$
$R_{1/2}$
$R_{3/2}$
$R_{4/2}$
$R_{5/2}$
$R_{6/2}$
+
+
$-R_{1/2}$
$-L_{2/1}$
L
R
cross-talk canceller
directionnal encoder
2-channel loudspeaker decoder

FIG. 11


directionnal encoder
4-channel loudspeaker decoder
in1
g_1
g_2
g_3
g_4
g_5
g_6
MIX
0.7
$L_{2/1}$
$L_{3/1}$
$L_{3/5}$
$L_{4/5}$
$L_{6/1}$
$L_{6/5}$
$R_{1/2}$
$R_{3/2}$
$R_{3/4}$
$R_{5/4}$
$R_{6/2}$
$R_{6/4}$
cross-talk canceller
$-R_{1/2}$
$-L_{2/1}$
$-R_{5/4}$
$-L_{4/5}$
LF
RF
(front)
LB
RB
(back)

Please type a plus sign (+) inside this box → [+]

PTO/SB/01 (10-00)

Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)

☐ Declaration Submitted With Initial Filing OR ☒ Declaration Submitted after Initial Filing (surcharge (37 CFR 1.16 (e)) required)

Attorney Docket Number	017002-012720US
First Named Inventor	Jot, Jean-Marc
COMPLETE IF KNOWN	
Application Number	09/806,193
Filing Date	
Group Art Unit	
Examiner Name	

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

METHOD AND APPARATUS FOR THREE-DIMENSIONAL AUDIO DISPLAY

the specification of which *(Title of the Invention)*

☐ is attached hereto

OR

☒ was filed on (MM/DD/YYYY) 09/24/1999 as United States Application Number or PCT International Application Number PCT/US99/22259 and was amended on (MM/DD/YYYY) ______ (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims as amended specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or of any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY) Country	Priority Not Claimed	Certified Copy Attached? YES	NO
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐

☐ Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto:

I hereby claim the benefit under 35 U.S.C. 119(e) of any United States provisional application(s) listed below.

Application Number(s)	Filing Date (MM/DD/YYYY)
60/101,884	September 24, 1998

☐ Additional provisional application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

Please type a plus sign (+) inside this box → [+]

PTO/SB/01 (10-00)
Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION — Utility or Design Patent Application

Direct all correspondence to:	☒ Customer Number or Bar Code Label	20350	OR	☐ Correspondence address below
Name				
Address				
Address				
City		State		ZIP
Country	Telephone			Fax

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

NAME OF SOLE OR FIRST INVENTOR: ☐ A petition has been filed for this unsigned inventor

Given Name: Jean-Marc		Family Name or Surname: Jot	
Inventor's Signature:			Date: 16 August 2001
Residence: City: Aptos CA	State: CA	Country: US	Citizenship: France
Mailing Address: ~~515 Townsend Drive~~ 327 DORIS AVENUE			
Mailing Address			
City: Aptos	State: CA	ZIP: 95003	Country: US

NAME OF SECOND INVENTOR: ☐ A petition has been filed for this unsigned inventor

Given Name: Scott		Family Name or Surname: Wardle	
Inventor's Signature: Scott Wardle		Date: 7 September 2001	
Residence: City: Santa Cruz CA	State: CA	Country: US	Citizenship: US
Mailing Address: 823 Riverside Avenue			
Mailing Address			
City: Santa Cruz	State: CA	ZIP: 95060	Country: US

☐ Additional Inventors are being named on the _____ supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto.

SF 1256783 v1

09/806193

Please type a plus sign (+) inside this box → [+]

Rec'd PCT/PTO 09 JAN 2002

PTO/SB/81 (02-01)
Approved for use through 10/31/2002. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

POWER OF ATTORNEY OR AUTHORIZATION OF AGENT

Application Number	09/806,193
Filing Date	
First Named Inventor	Jot, Jean-Marc
Title	METHOD AND APPARATUS FOR THREE-DIMENSIONAL AUDIO DISPLAY
Group Art Unit	
Examiner Name	
Attorney Docket Number	017002-012720US

I hereby appoint:

☒ Practitioners at Customer Number 20350 → *Place Customer Number Bar Code Label here*

OR

☐ Practitioner(s) named below:

Name	Registration Number

as my/our attorney(s) or agent(s) to prosecute the application identified above, and to transact all business in the United States Patent and Trademark Office connected therewith.

Please change the correspondence address for the above-identified application to:

☒ The above-mentioned Customer Number.

OR

☐ Practitioners at Customer Number

☐ Firm *or* Individual Name					
Address					
Address					
City		State		ZIP	
Country					
Telephone		Fax			

I am the:

☐ Applicant/Inventor.

☒ Assignee of record of the entire interest. See 37 CFR 3.71. **(Creative Technology, Ltd.)**

Statement under 37 CFR 3.73(b) is enclosed. (Form PTO/SB/96).

SIGNATURE of Applicant or Assignee of Record

Name	TERRENCE CARROLL
Signature	
Date	September 13, 2001

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of ____ forms are submitted.

SF 1256820 v1